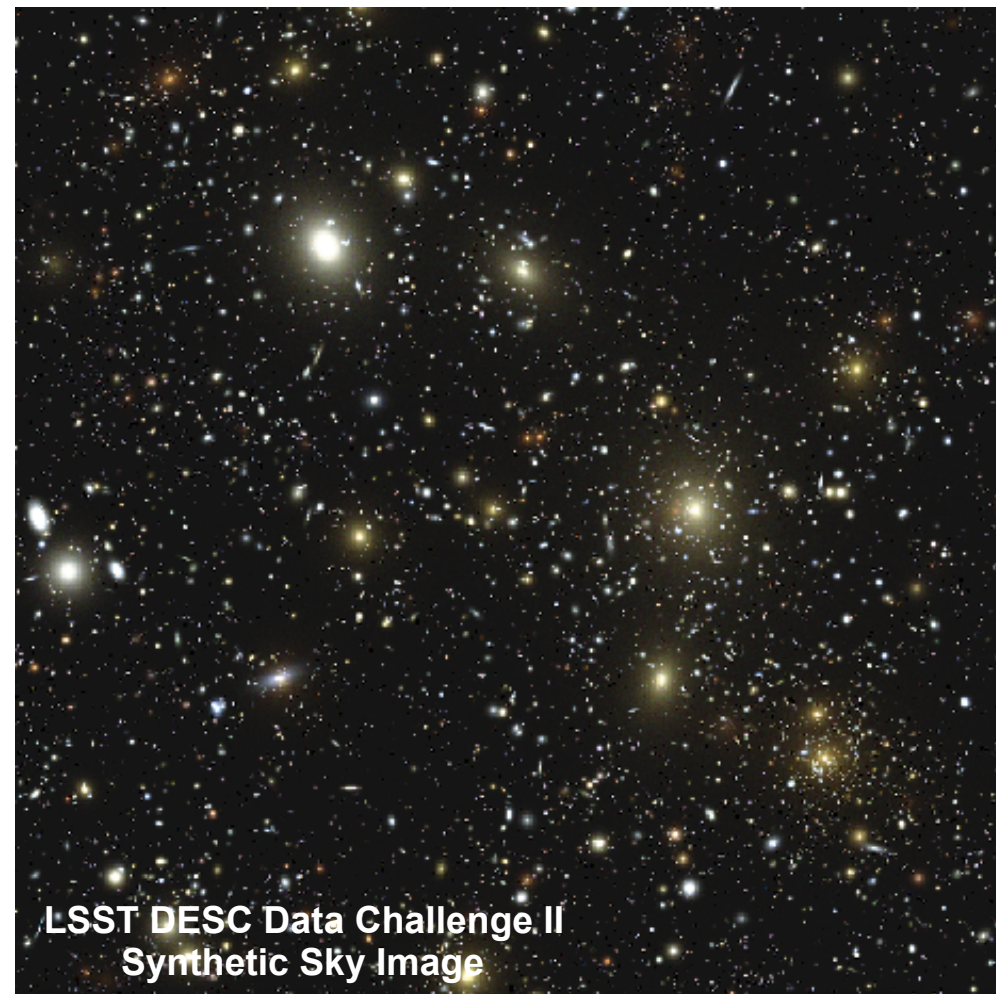


# ECP Computational Cosmology Drivers for Intelligent Data Movement

Salman Habib  
(for ExaSky)  
Argonne National Laboratory

## Science workflows:

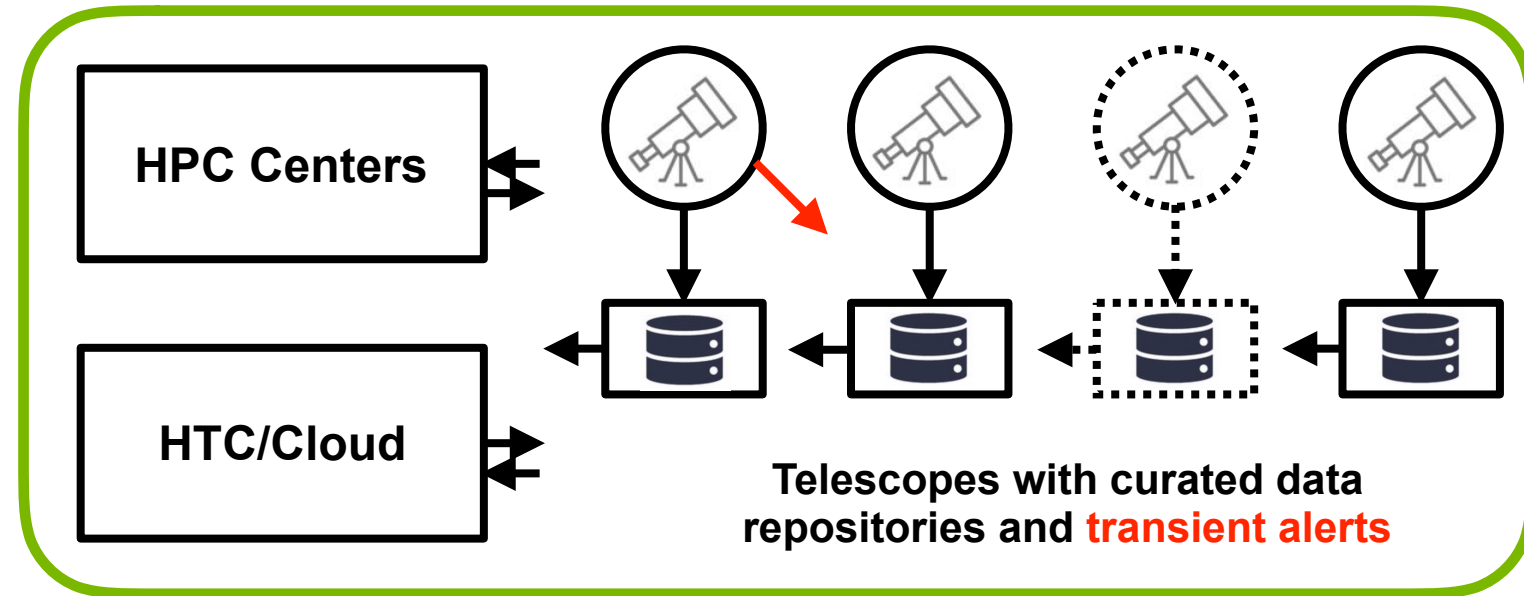
- Time-varying and (mostly) static datasets
- Melding simulation-generated and observational data sets with AI/ML-aided analysis tools
- Distributed data sources (telescopes), data repositories, and distributed computing resources
- Complex problem involving multiple agencies, data and compute hubs, heterogeneous user community, at-scale distributed workflow management



# Data Connectivity in the Cosmology Environment

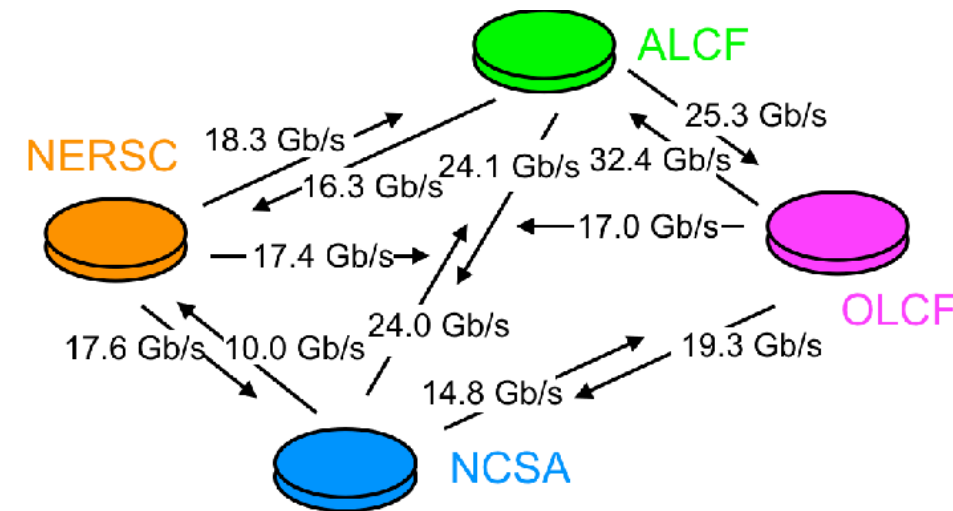
## Workflow Elements

- Global orchestration for production-scale activities
- Local workflow management
- Allowance for individual scientist, small teams, and large collaboration driven projects
- Nuts and bolts — data rights, authentication, cross-site resource management
- Automated data transfers driven by workflows
- Computations at all scales, from small 'event' sizes to 'all-sky' analyses



## Some Numbers

- **Data:** sizes at the 100's of PB
- **Transfer:** up to ~PB/day
- **Compute:** ~10% of all US computing centers



Petascale data transfer project with ESnet (via Globus)